

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2003, 15:04:31 ; Search time 23 Seconds
(without alignments)
331.811 Million cell updates/sec

Title: AAK91826

Perfect score: 965
Sequence: 1 MKRGPSLRGRDAPAPPCV.....ATELGSTELVTKIAGPEQO 184

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	965	100.0	184	T13C_HUMAN	Q96TJ3 homo sapien
2	410.5	42.5	175	T13C_MOUSE	Q968D0 mus musculu
3	124	12.8	439	XP2_XENLA	P17437 xenopus lae
4	120.5	12.5	1083	T2D3_HUMAN	O00268 homo sapien
5	116.5	12.1	676	ICP0_HSVBJ	P29128 bovine herp
6	109	11.3	325	DBP_HUMAN	O10586 homo sapien
7	107	11.1	2142	BAT2_HUMAN	P48634 homo sapien
8	106.5	11.0	356	HXB2_HUMAN	P14652 homo sapien
9	103	10.7	505	WAS1_BOVIN	Q95107 bos taurus
10	102.5	10.6	676	ICP0_HSVBK	P29836 bovine herp
11	102.5	10.6	725	AREA_PENCH	O01582 penicillium
12	102	10.6	317	YQ35_CAEEL	O09456 caenorhabd
13	102	10.6	325	DBP_MOUSE	O60925 mus musculu
14	102	10.6	325	DBP_RAT	P16443 rattus norv
15	99.5	10.3	503	IRF7_HUMAN	Q92985 homo sapien
16	99	10.3	1487	ICP4_HSVB	P28925 equine herp
17	98	10.2	684	CA39_HUMAN	P17473 equine herp
18	98	10.2	1487	ICP4_HSVK	P17473 equine herp
19	97.5	10.1	384	VASP_CNFA	P50551 canis fami
20	97.5	10.1	915	AL80_RAT	O05190 pseudomom
21	97	10.1	353	LICH_PSEGL	O04808 mus musculu
22	97	10.1	1229	N121_HUMAN	Q97293 homo sapien
23	96.5	10.0	1255	D1A1_MOUSE	O08608 mus musculu
24	96	9.9	3164	TEGU_HSV11	O97566 homo sapien
25	96	9.9	3164	TEGU_HSV11	O97566 homo sapien
26	95.5	9.9	129	FN14_HUMAN	Q91820 herpes simp
27	95.5	9.9	316	CC07_CAEEL	P18832 caenorhabd
28	95	9.8	412	CBX6_HUMAN	O95503 caenorhabd
29	95	9.8	625	DUS8_HUMAN	O13302 homo sapien
30	95	9.8	722	Z219_HUMAN	O92974 homo sapien
31	95	9.8	776	SM4F_RAT	Q92143 rattus norv
32	95	9.8	1248	D1A1_HUMAN	O60610 homo sapien
33	95	9.8	1509	GSRI_HUMAN	O9azm4 homo sapien

34	94.5	9.8	374	1	PSPD_MOUSE	P50404 mus musculu
35	94.5	9.8	449	1	ADP_BRANA	P40603 brassica na
36	94.5	9.8	1838	1	CA15_HUMAN	P20908 homo sapien
37	94	9.7	263	1	ICP3_HSV1F	P08353 herpes simp
38	94	9.7	654	1	SPH2_HUMAN	O91840 homo sapien
39	94	9.7	777	1	SM4F_MOUSE	O92123 mus musculu
40	94	9.7	806	1	MK07_MOUSE	Q94949 mus musculu
41	93.5	9.7	421	1	FXJ1_HUMAN	Q92949 homo sapien
42	93.5	9.7	815	1	MK07_HUMAN	O13164 homo sapien
43	93.5	9.7	1290	1	PER1_HUMAN	O15534 homo sapien
44	93	9.6	184	1	TR17_HUMAN	O02223 homo sapien
45	93	9.6	185	1	TR17_MOUSE	O88472 mus musculu

ALIGNMENTS

```

RESULT 1
ID T13C_HUMAN STANDARD; PRT; 184 AA.
AC Q96RJ3;
DT 15-JUN-2002 (Rel. 41, last, sequence update)
DT 15-JUN-2002 (Rel. 41, last, annotation update)
DE Tumor necrosis factor receptor superfamily member 13C (B cell-
DE activating factor receptor) (BAFF receptor) (BAFF-R) (Blys receptor
DE 3).
GN TNFRSF13C OR BAFFR OR BR3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISORFMS 1 AND 2).
RC TISSUE=B-cell lymphoma;
RA MEDLINE=21442025; PubMed=11509692;
RA Thompson J.S., Bixler S.A., Qian F., Vora K., Scott M.L.,
RA Cachero T.G., Hession C., Schneider P., Sizig I.D., Mullen C.,
RA Strach K., Zafari M., Benjamin C.D., Teichopp J., Browning J.L.,
RA Ambrose C.; a newly identified TNF receptor that specifically interacts
RA "BAFF-R,"
RA with BAFF,"
RA Science 293:2108-2111 (2001).
RL [2]
RP FUNCTION.
RA MEDLINE=21475520; PubMed=11591325;
RA Yan W., Brady J.R., Chan B., Lee W.P., Hsu B., Hartless S.M.,
RA Canero M.P., Grewal I.S., Dixit V.M.;
RA "Identification of a novel receptor for B lymphocyte stimulator that
RA is mutated in a mouse strain with severe B cell deficiency,"
RA Curr. Biol. 11:1547-1552 (2001).
RL CC - FUNCTION: B-cell receptor specific for TNFRSF1B/TNLF1/BAFF/Blys.
CC Promotes the survival of mature B-cells and the B-cell response.
CC - SUBCELLULAR LOCATION: Type III membrane protein (Probable).
CC - ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be
CC produced by alternative splicing.
CC - TISSUE SPECIFICITY: Highly expressed in spleen and lymph node, and
CC in resting CD4+ T-cells, in thymus and peripheral blood leukocytes.
CC - SIMILARITY: CONTAINS 1 TNFR-CYS REPEAT.
CC - This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
DR EMBL; AF373846; AAK91826.1;
DR GenBank; HGNC:17755; TNFRSF13C.
DR MIM; 606269;
DR PROSITE; PS00652; TNFR_NGFR_1; FALSE_NEG.

```

DR PROSITE; PSS0050; TNFR NGFR 2; FALSE NEG.
 KM Receptor; Immune response; Signal-anchor; Transmembrane;
 KW Alternative splicing 78
 FT DOMAIN 1
 FT TRANSMEM 79 99
 FT SIGNAL-ANCHOR (POTENTIAL)
 FT DOMAIN 100 184
 FT REPEAT 18 35
 FT DISULFID 19 32
 FT DISULFID 24 35
 FT VARSPLIC 143 143
 SQ SEQUENCE 184 AA; 18863 MW; F2BFB9809A27138 CRC64;
 P -> PA (IN ISOFORM 2).
 Query Match 100.0%; Score 965; DB 1; Length 184;
 Best local similarity 100.0%; Pred. No. 1e-56;
 Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MRGPRSLGRDAPAPPCVPAECFDLVNHCYACGLLTPRDKPAGASSPAPRTALQPO 60
 DB 1 MRGPRSLGRDAPAPPCVPAECFDLVNHCYACGLLTPRDKPAGASSPAPRTALQPO 60
 DB 61 ESAGAGAGAAALPLGLLTPAPALGLAVLALVWGLVSWRRORRLGASAAAPDDG 120
 DB 61 ESAGAGAGAAALPLGLLTPAPALGLAVLALVWGLVSWRRORRLGASAAAPDDG 120
 QY 121 KDAPELDKVIILSPGISDAPAPPPGEGPTPPGHSVPVPAELGSTEIVTKTAG 180
 DB 121 KDAPELDKVIILSPGISDAPAPPPGEGPTPPGHSVPVPAELGSTEIVTKTAG 180
 QY 181 PEGQ 184
 DB 181 PEGQ 184
 DB 181 PEGQ 184
 RESULT 2
 ID T13C.MOUSE STANDARD; PRT; 175 AA.
 AC Q9DBD0;
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Tumor necrosis factor receptor superfamily member 13C (B cell-
 activating factor receptor) (BAFF receptor) (BAFF-R) (Blys receptor
 3) (B-cell maturation defect).
 GN TNFRSF13C OR BAFFR OR BCMD OR BR3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCB1_TaxID=10090;
 (1)
 RC SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC STRAIN=BA1B/C; TISSUE=B-cell lymphoma;
 RX MEDLINE=21442025; PubMed=11509692;
 RA Thompson J.S., Bixler S.A., Qian F., Vora K., Scott M.L.,
 RA Cachero T.G., Hession C., Schneider P., Sizling J.D., Millen C.,
 RA Strauch K., Zafari M., Benjamin C.D., Tschopp J., Browning J.L.,
 RA Ambrose C.;
 RT "BAFF-R, a newly identified TNF receptor that specifically interacts
 with BAFF.";
 RT Science 293:2108-2111(2001).
 RN (12)
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND DISEASE.
 RC STRAIN=A/J;
 RX MEDLINE=21475500; PubMed=11591325;
 RA Yan M., Brady J.R., Chan B., Lee W.P., Hsu B., Harless S.M.,
 RA Canco M.P., Grewal I.S., Dixit V.M.;
 RT "Identification of a novel receptor for B lymphocyte stimulator that
 is mutated in a mouse strain with severe B cell deficiency.";
 RT Curr. Biol. 11:1547-1552(2001).
 RN (3)
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC STRAIN=C57BL/60; TISSUE=Small intestine;
 RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kanakawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochava H.,
 RA Kuehl P., Lewis S., Matsumoto Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guernicich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombereite P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlschki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN (4)
 RP FUNCTION.
 RX MEDLINE=21614654; PubMed=11747827;
 RA Harless S.M., Lentz V.M., Sah A.P., Hsu B.L., Clise-Dwyer K.,
 RA Hilbert D.M., Hayes C.E., Cancro M.P.;
 RT "Competition for Blys-mediated signaling through Bcmd/BR3 regulates
 peripheral B lymphocyte numbers.";
 RL Curr. Biol. 11:1986-1989(2001).
 CC -1- FUNCTION: B-cell receptor specific for TNFRSF1B/TNFI/BAFF/Blys.
 CC Promotes the survival of mature B-cells and the B-cell response.
 CC -1- SUBCELLULAR LOCATION: Type III membrane protein (probable).
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are
 CC produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Highly expressed in spleen and testis;
 CC detected at lower levels in lung and thymus.
 CC -1- DISEASE: Defects in TNFRSF13C are a cause of severe B-cell
 CC deficiency. B-cell deficient strain A/WySn has a 4.7 kb insertion
 CC in the BAFFR gene leading to an altered C-terminal. The mutant RNA
 CC is not detectable. B-cell lymphopoeisis is normal, but the life
 CC span of peripheral B-cells is much reduced.
 CC -1- SIMILARITY: CONTAINS 1 TNFR-CYS REPEAT.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC DR EMBL; AF373847; AAK91827.1; -;
 CC DR EMBL; AK008142; BAB25490.1; -;
 CC MGD; MGI:1919299; Tnfrsf13c.
 KW Receptor; Immune response; Signal-anchor; Transmembrane; Glycoprotein;
 KW Alternative splicing.
 FT DOMAIN 1
 FT TRANSMEM 72 92
 FT SIGNAL-ANCHOR (POTENTIAL)
 FT DOMAIN 100 184
 FT REPEAT 18 35
 FT DISULFID 19 32
 FT DISULFID 27 38
 FT CARBOHYD 23 23
 FT VARSPLIC 133 143
 SQ SEQUENCE 175 AA; 18798 MW; 28BC7C1A02FB7EF CRC64;
 P -> PA (IN ISOFORM 2).
 Query Match 42.5%; Score 410.5; DB 1; Length 175;
 Best local similarity 56.1%; Pred. No. 1.9e-21;
 Matches 101; Conservative 9; Mismatches 55; Indels 15; Gaps 6;
 QY 6 RSLGRDAPAPPCVPAECFDLVNHCYACGLLTPRDKPAGASSPAPRTALQPOESVVA 65
 DB 9 RSLGRDAPAPPCVPAECFDLVNHCYACGLLTPRDKPAGASSPAPRTALQPOESVVA 62

QY	66	GACGAAALPELPGLLGCAPALLGLAVLAIV-LVGLVSWRRRRRLRGASSAAPQDCKXA-123
Db	63	--GSLARDVALVGVAPALLGLLIALTLVGLVSLVSWMRQ-QLRITAS-----PDISEGVQ115
OY	124	PEPLDKVIIISPGISDAPAPAMPPEPGEPTTPPEGHSVPVPATELIGSTELVTTKTAGEQ183
Db	116	QESLENNVFPVPSSETPHAGAPTPMPPEKEDADSDALPRHSVYPVATELIGSTELVTTKTAGEQ175
RESULT 3		
XP2_XENIA	STANDARD;	PRT; 439 AA.
ID	XP2_XENIA	
AC	P17437; Q08944;	
DT	01-AUG-1990 (Rel. 15, Created)	
DT	01-FEB-1995 (Rel. 31, Last sequence update)	
DT	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	SKIN secretory protein XP2 precursor (APBG protein).	
	Xenopus laevis (African Clawed frog).	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;	
	Xenopodinae; Xenopus.	
OC	NCBI_TaxID=8355;	
RP	(1)	
RP	SEQUENCE OF 1-25 AND 344-439 FROM N.A. (ISOFORM 2).	
RX	MEDLINE=92332564; PubMed=1629230;	
RA	Hauser F., Roeben C., Hoffmann W.;	
RT	"xp2, a new member of the p-dominant peptide family of potential growth	
RL	factors, is synthesized in Xenopus laevis skin.";	
RL	J. Biol. Chem. 267:14451-14455(1992).	
RN	(2)	
RP	SEQUENCE OF 3-439 FROM N.A. (ISOFORM 1).	
RC	TISSUE=Skin;	
RX	MEDLINE=90127399; PubMed=2292893;	
RA	Gmachl M., Berger H., Thalhammer U., Kreil G.;	
RT	"dermal glands of Xenopus laevis contain a polypeptide with a highly	
RL	repetitive amino acid sequence.";	
RL	FEBS Lett. 260:145-148(1990).	
CC	-1- FUNCTION: MAY ACT AS A GROWTH FACTOR IN THE GERMINAL LAYER OF THE	
CC	EPIDERMIS. MAY ALSO BE INVOLVED IN GROWTH OF REGENERATING GLANDS	
CC	AND IN PROTECTION OF THE SKIN FROM THE EXTERNAL ENVIRONMENT.	
CC	-1- SUBCELLULAR LOCATION: Secreted.	
CC	-1- ALTERNATIVE PRODUCTS: 2 isoforms; 1/APBG (shown here) and 2/XP2;	
CC	are produced by alternative splicing.	
CC	-1- TISSUE SPECIFICITY: SKIN.	
CC	-1- SIMILARITY: CONTAINS 2 P-TYPE (TREPOLI) DOMAINS.	
CC	-1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 392	
CC	ONWARD AND IS SHORTER (418 AA) DUE TO A FRAMESHIFT.	

CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce)	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; M90095; AAA50001.1; -;	
DR	EMBL; X51394; CA35759.1; ALT_FRAME.	
DR	PIR; S07498; SKXLAG.	
DR	PIR; A37331; A37331.	
DR	HSSP; P04155; 1P62.	
DR	InterPro; IPR000519; P_trefo1l.	
DR	Pfam; PF00088; trefo1l.2.	
DR	PRINTS; PR00680; PTREPOLI.	
DR	SMART; SM00018; P_2.	
DR	PROSITE; PS00025; P_TREPOLI_2.	
RV	Signal; Growth factor; Alternative splicing; Amphibian skin; Repeat.	
FT	SIGNAL	1 22
FT	CHAIN	23 439
FT	MOD RES	26 23
FT	DOMAIN	33 343
FT	DOMAIN	350 391
	P-TYPE 1.	

FT	DORAIN	397	438	P-TYPE 2.
FT	DISULFID	351	377	BY SIMILARITY.
FT	DISULFID	361	376	BY SIMILARITY.
FT	DISULFID	371	388	BY SIMILARITY.
FT	DISULFID	398	424	BY SIMILARITY.
FT	DISULFID	408	423	BY SIMILARITY.
FT	DISULFID	418	435	BY SIMILARITY.
FT	VARSPLIC	26	343	MISSING (IN ISOFORM 2).
FT	CONFLICT	3	3	H -> S (IN REF. 2).
FT	CONFLICT	18	18	C -> W (IN REF. 2).
SO	SEQUENCE	439 AA;	41173 MM;	38C4A4B57CBAE778 CRC64;
Query March 12.8% Score 124; DB 1; Length 439;				
Best local similarity 27.5% Pred. No. 0.08;				
Matches 46; Conservative 10; Mismatches 57; Indels 54; Gaps 7				
Qy	12	DAPAPTPC-----VPAPCFDLVNRHCVCAGLRTTPRKPAGSAPPARTLQPOESVG	64	
Db	155	EAPAPAPAEVAPAPAPAE-----EAPAPAPAEAGAPAPAPAE	200	
Qy	65	AGAGEALPLPGLLFGAPALLGLALVLVLVGLVSWRRRRRLRGASSAEP-DGDKDA	123	
Db	201	PAEEAPAPAP-----APA-----EGEAPAPAPAE	230	
Qy	124	PEPDL-KTILSPGSDTAPAPPPGSDPTTPPGHSPVPATELG	169	
Db	231	PAPAPGEAPAPAPAEAPAPAPAE-----PAPAPAPAPAEAPAPAPAE	276	
RESULT 4				
ID	T2D3_HUMAN	STANDARD;	PRT;	1083 AA.
AC	000268;	Q99721; Q9BX42; Q9BR40;		
DT	15-JUL-1998	(Rel. 36, Created)		
DT	15-JUL-1998	(Rel. 36, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		
DE	Transcription initiation factor TFIID 135 kDa subunit (TAFII-135)			
DE	(TAFII135) (TAFII-130) (TAFII130).			
GN	TAF4 OR TAF4A OR TAF2C1 OR TAF2C OR TAFII135 OR TAFII130.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxId=9606;			
RN	[1]			
RN	SEQUENCE FROM N.A.			
RX	MEDLINE=9735072; PubMed=9192867;			
RA	Mengus G., May M., Carre L., Chabon P., Davidson I.;			
RT	"Human TAF(II)135 potentiates transcriptional activation by the AP-2s			
RT	of the retinoic acid, vitamin D3, and thyroid hormone receptors in			
RT	mammalian cells.";			
RL	Genes Dev. 11:1381-1395(1997).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=2638749; PubMed=11780052;			
RA	DeJouvas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,			
RA	Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,			
RA	Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,			
RA	Beesley O.P., Bird C.P., Blakey S.E., Bridgman A.M., Brown A.J.,			
RA	Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,			
RA	Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,			
RA	Clagg S., Cobley V.E., Collier R.E., Connor R.E., Coty N.R.,			
RA	Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,			
RA	Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,			
RA	Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,			
RA	Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,			
RA	Huckle B., Hunt A.R., Hunt S.B., Jeksch K., Johnson C.M., Johnson D.,			
RA	Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,			
RA	Levansthal M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,			
RA	Marsh V.L., Martin S.L., McConachie I.J., McMay K., Murray A.A.,			
RA	Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,			
RA	Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,			
RA	Phillimore B.J.C.T., Pratchinagam S.R., Plumb R.W., Ramsay H.,			
RA	Rice C.M., Ross M.T., Scott C.E., Senta H.K., Showkeen R., Sims S.,			

RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.M., Symamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin R., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilmink L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.,
 RT "The DNA sequence and comparative analysis of human chromosome 20.",
 RL Nature 414:865-871(2001).
 RN (3)
 RN SEQUENCE OF 105-1083 FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=97098442; PubMed=8942982;
 RA Tanase N., Saluja D., Vaseallo M.F., Chen J.-L., Admon A.,
 RT "Molecular cloning and analysis of two subunits of the human TFIID
 RT complex: hTAF1130 and hTAF1100."
 RL Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616(1996).
 CC -1- FUNCTION: MAKES PART OF TFIID IS A MULTIMERIC PROTEIN COMPLEX THAT
 CC PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER ACTIVATION
 CC ACTIVATORS AND REPRESSORS. POTENTIATES TRANSCRIPTIONAL ACTIVATION
 CC BY THE AP-25 OF THE RETINOIC ACID, VITAMIN D3 AND THYROID HORMONE.
 CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
 CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).
 CC -----
 CC DR EMBL: Y11354; CAA72189.1; -
 CC DR EMBL: AL137077; CAC36006.1; -
 CC DR EMBL: AL109911; CAC2312.2; -
 CC DR EMBL: U75308; AAC50901.1; -
 CC DR TRANSFAC: T02328; -
 CC DR GeneW: HGNC:11537; TAF4.
 CC DR MIM: 601796; -
 CC DR InterPro: IPR003894; TAF_hom.
 CC DR SMART: SM00549; TAFH, 1.
 CC DR SMART: SM00549; TAFH, 1.
 CC KW Transcription regulation; Nuclear protein.
 CC FT DOMAIN 39 42 POLY-HIS.
 CC FT DOMAIN 52 57 POLY-ALA.
 CC FT DOMAIN 98 101 POLY-GLY.
 CC FT DOMAIN 142 148 POLY-ALA.
 CC FT DOMAIN 268 275 POLY-PRO.
 CC FT DOMAIN 331 337 POLY-ALA.
 CC FT DOMAIN 680 683 POLY-PRO.
 CC FT DOMAIN 808 813 POLY-ALA.
 CC FT DOMAIN 828 831 POLY-ASP.
 CC FT DOMAIN 105 117 PGPPRRPLVPA -> GRGLLQRRGRES
 CC (IN REF. 3)
 CC FT CONFLICT 136 136 A -> S (IN REF. 2).
 CC FT CONFLICT 185 185 G -> GPG (IN REF. 2).
 CC FT CONFLICT 233 264 MISSING (IN REF. 3).
 CC FT CONFLICT 293 293 P -> L (IN REF. 3).
 CC SO SEQUENCE 1083 AA; 109943 MW; A6453827572A0752 CRC64;
 Query Match 12.5%; Score 120.5; DB 1; Length 1083;
 Best Local Similarity 27.4%; Pred. No. 0.31; Indels 63; Gaps 9;
 Matches 63; Conservative 11; Mismatches 93;
 QY 2 RRGPRRLRGDAPAPPCVPA-----ECFDLVRHCVCAGLRTPPKPDAGAS 49
 DB 104 RRGPRRLRVPA-GRAPPAALRRPPBSAGACAPVPAALVAAAG---PEPADGPA 158
 QY 50 SPAPRTALQPOESVGAGGAPALPLPG---LFGAPALL----- 85
 DB 159 KPAGPALAARAGPGPGPGPGKPAAGPAGAGTGLNSALLNSHHAAPAVSLVNN 218
 QY 86 GLALVLLVVLGVSRRRRLRGASAAEPDGDADPEPLDKVITLLSGISDAPAPM 145

DB 219 GPALLPLPKPAPAGTGTTPPEVGAAPAPAP---AAPSPAPAPAPAP---AAAPP 270
 QY 146 PPGEDPGTTPPGH-----SYVPAT-----ELGSTELVTTKTGP 181
 DB 271 PEPAPRTALRRPGRHAGPPTAAPVPPPAAGSAGAPAPAPAGPG 320
 RESULT 5
 ICPO_HSVBJ STANDARD; PRT; 676 AA.
 AC P29128;
 DT 01-DEC-1992 (rel. 24, Created)
 DT 01-DEC-1992 (rel. 24, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Trans-acting transcriptional protein ICPO (p135 protein) (IER
 DE 2.9/IER.6).
 GN BICPO.
 OS Bovine herpesvirus type 1 (strain Jura).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicellovirinae.
 OX NCBI_Taxid=31518;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92219360; PubMed=1313901;
 RA Wirth U.V., Fraefel C., Vogt B., Vlcek C., Paces V., Schwyzer M.;
 RT "Immediate-early RNA 2.9 and early RNA 2.6 of bovine herpesvirus 1
 RT are 3' coterminal and encode a putative zinc finger transactivator
 RT protein."
 RL J. Virol. 66:2763-2772(1992).
 CC -1- PTM: THE STRONGLY ACIDIC REGION MIGHT SERVE AS A TRANSCRIPTIONAL
 CC ACTIVATION DOMAIN, POSSIBLY REGULATED THROUGH PHOSPHORYLATION BY
 CC CASEIN KINASE II.
 CC -1- SIMILARITY: TO OTHER HERPESVIRUSES ICPO PROTEIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).
 CC -----
 CC DR EMBL: M84465; AAA46062.1; -
 CC DR EMBL: AJ004801; CAA06138.1; -
 CC DR PIR: B38209; EDBE23.
 CC DR HSSP: P28990; 1CHC.
 CC DR InterPro: IPR001841; Znf ring.
 CC DR Pfam: PF00097; zf-C3HC4_1.
 CC DR SMART: SM00184; RING, 1.
 CC DR PROSITE: PS00518; ZF_RING_1; 1.
 CC DR PROSITE: PS00589; ZF_RING_2; 1.
 CC KW Transcription regulation; Trans-acting factor; Activator; Zinc-finger;
 CC KW DNA-binding; Early protein; Repressor; Phosphorylation.
 CC FT ZN_FING 13 52 RING-TYPE.
 CC FT DOMAIN 284 331 ASP/GLU-RICH (ACIDIC).
 CC SO SEQUENCE 676 AA; 67879 MW; 11806BA4E5C4B71 CRC64;
 Query Match 12.1%; Score 116.5; DB 1; Length 676;
 Best Local Similarity 30.4%; Pred. No. 0.38; Indels 51; Gaps 11;
 Matches 62; Conservative 13; Mismatches 78;
 QY 7 SLRGDAPAPPCVPAECFDLVRHCVCAG-----LARTPRKP 45
 DB 353 STRGRQTRAVQPARPS-----LARR--FCGRAAAVASPPSSRGGRDRPLPAPRAP 405
 QY 46 AG---ASSPAPRTALQPOESVGAGGAPALPLPGLLFGAPALLGLALVAVLVG---LV 99
 DB 406 AAQARACSPBPR-----EEGRAGLGVAAGETAGAGASSEEGRRARRLLGEAPPRVQ 460
 QY 100 SWRRQRRL-RGASAAEPDGDADPEPLDKVITLLSGISDAPAPMPPPPEDPGTTP 157
 DB 461 ARRRRTTDLRAPTPAPAP---APAPAPITVITD---ANAPAPADDPAPAPAGPASA 514

QY 158 GHSVPATGTELVYTKTAP 181
 DB 515 GAQIGTPA-----AAAATTAATAAP 534

RESULT 6
 DBP_HUMAN STANDARD; PRT; 325 AA.
 AC 010586;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE D-site-binding protein (Albumin D box-binding protein) (TAXREB302).
 GN DBP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=95137580; PubMed=7835883;
 RA Khatib Z.A., Inaba T., Valentine M., Look A.T.;
 RT "Chromosomal localization and cDNA cloning of the human DBP and TEF
 genes.";
 RL Genomics 23:344-351(1994).
 [2]
 SEQUENCE FROM N.A.
 MEDLINE=96374625; PubMed=8786133;
 RA Shuttler G., Glasco T., Kang X., Korneluk R., Mueller C.R.;
 RT "Genomic structure of the human D-site binding protein (DBP). gene.";
 RL Genomics 34:334-339(1996).
 [3]
 SEQUENCE FROM N.A.
 TISSUE=Brain;
 MEDLINE=97264341; PubMed=9110174;
 RA Yu W., Andersson B., Worley K.C., Muzny D.M., Ding Y., Liu W.,
 RA Ricafrente J.Y., Wentland M.A., Lennon G., Gibbs R.A.;
 RT "Large-scale concatenation cDNA sequencing.";
 RL Genome Res. 7:353-358(1997).
 [4]
 SEQUENCE OF 132-325 FROM N.A.
 MEDLINE=93346252; PubMed=8482542;
 RA Nyunoya H., Morita T., Sato T., Honma S., Tsujimoto A.,
 RA Shimotohno K.;
 RT "Cloning of a cDNA encoding a DNA-binding protein TAXREB302 that is
 specific for the tax-responsive enhancer of HTLV-I.";
 RL Gene 126:251-255(1993).
 [5]
 REVISIONS.
 RA Nyunoya H.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 [6]
 REVIEW.
 MEDLINE=9439627; PubMed=10508692;
 RA Brown S.A., Schibler U.;
 RT "The ins and outs of circadian timekeeping.";
 RL Curr. Opin. Genet. Dev. 9:588-594(1999).
 [7]
 FUNCTION: THIS TRANSCRIPTIONAL ACTIVATOR RECOGNIZES AND BINDS TO
 THE SEQUENCE 5'-ATTAGTAA-3' FOUND IN THE PROMOTER OF GENES SUCH
 AS ALBUMIN, CYP2A4 AND CYP2A5. IT IS NOT ESSENTIAL FOR CIRCADIAN
 RHYTHM GENERATION, BUT MODULATES IMPORTANT CLOCK OUTPUT GENES. MAY
 BE A DIRECT TARGET FOR REGULATION BY THE CIRCADIAN PACEMAKER
 COMPONENT CLOCK. MAY AFFECT CIRCADIAN PERIOD AND SLEEP REGULATION.
 [8]
 SUBUNIT: BINDS DNA AS A HOMODIMER OR A HETERODIMER. CAN FORM A
 HETERODIMER WITH TEF.
 [9]
 SUBCELLULAR LOCATION: Nuclear.
 [10]
 TISSUE SPECIFICITY: UBICUITOUSLY EXPRESSED. EXPRESSED IN THE
 SUPRACHIASMATIC NUCLEI (SCN) AND IN MOST PERIPHERAL TISSUES, WITH
 A STRONG CIRCADIAN RHYTHMICITY.
 [11]
 SIMILARITY: BELONGS TO THE BZIP FAMILY. PAR SUBFAMILY.
 [12]
 This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, U06936; AAA81374.1; -
 CC EMBL, U48213; AAB18668.1; -
 CC EMBL, U48212; AAB18668.1; JOINED.
 CC EMBL, U79283; AAB50219.1; -
 CC EMBL, D28468; BAA05833.1; -
 CC TRANSFAC; T04875; -
 CC DR Genew; HGNC:2697; DBP.
 CC MIM: 124097; -
 CC DR InterPro; IPR004827; TF_BZIP.
 CC DR SMART; SM00338; BRLZ.1.
 CC KW DNA-binding; Transcription regulation; Activator; Nuclear protein;
 CC Biological rhythms.
 CC FT DOMAIN 129 135 POLY-PRO.
 CC FT DOMAIN 188 239 PRO-RICH (PROLINE/ACIDIC REGION (PAR)).
 CC FT DNA_BIND 258 277 BASIC MOTIF.
 CC FT DOMAIN 250 304 LEUCINE-ZIPPER.
 CC FT CONFLICT 179 179 S -> T (IN REF. 1).
 CC FT CONFLICT 245 245 R -> K (IN REF. 4).
 CC SQ SEQUENCE 325 AA; 34349 MW; A6933CE21399ECF3 CRC64;
 Query Match 11.3%; Score 109; DB 1; Length 325;
 Best Local Similarity 27.1%; Pred.No. 0.62; 77; Indels 62; Gaps 9;
 Matches 56; Conservative 12; Mismatches 77; Indels 62; Gaps 9;
 QY 4 GPRSLRGDAPAPTCVPA-----ECFDL---LVKHCVCGLARTPPKPAQ 47
 DB 83 GGGSPRGPRGPVPAAGLAPLMERTLPQGVVEYVDLAFLEH---GL--PSPPPPG 136
 QY 48 ASSPARPALPOESVAGAGEAALP--LPGLFGAPALLGLALVLVLGVSWMRRQ 105
 DB 137 GSPSPPARPAPAPSPGSCGASPSRSPG---HAPRAALGT----- 177
 QY 106 RRLRGAASAEAPDGKADPEPLDKYIILSPGSDATAAMPWPGDPTTPPGHSVPVA 165
 DB 178 ----ASGHRAGLTNRDTPSPVD-----PTEVLTMTFEPDPAALASIPGHETPDR 226
 QY 166 TELGSTELY-----TTKTAGPEEQ 184
 DB 227 RHRFSEELKPOPIMKAKKIQVPEEQ 253

RESULT 7
 BAT2_HUMAN STANDARD; PRT; 2142 AA.
 AC P48634;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Large proline-rich protein Bat2 (HLA-B-associated transcript 2).
 GN BAT2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A.
 TISSUE=T-cell;
 MEDLINE=90192810; PubMed=2156268;
 RA Banerji J., Sands J., Strominger J.L., Spies T.;
 RT "A gene pair from the human major histocompatibility complex encodes
 large proline-rich proteins with multiple repeated motifs and a
 single ubiquitin-like domain.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:2374-2378(1990).
 [2]
 SEQUENCE OF 1-1860 FROM N.A.
 MEDLINE=93372029; PubMed=8499947;

RA Irie F.J.M., Bougueteloret L., Prieur S., Caterina D., Primes G.,
 RA Perrot V., Jurka J., Rodriguez-Tome P., Claverie J.-M., Dausset J.,
 RA Cohen D.;
 RT "Genes A1u clustering and a potential new member of the NF kappa B
 RT family within a 90 kilobase HLA class III segment.";
 RL Nat. Genet. 3:137-145(1993).
 CC -1- FUNCTION: UNKNOWN.
 CC -1- TISSUE SPECIFICITY: LIMITED TO CELL-LINES OF LEUKEMIC ORIGIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-8ib.ch/announce/>
 CC or send an email to license@1sb-8ib.ch).
 CC -----
 CC EMBL; M33509; AAA35585.1; -;
 DR EMBL; M33518; AAA35586.1; -;
 DR EMBL; M33512; AAA35586.1; JOINED.
 DR EMBL; Z15025; CAA78744.1; -;
 DR PIR; B35098; B35098.
 DR PIR; S36152; S36152.
 DR Genew; HGNC:13918; BAC72.
 DR MIM; 142580; -;
 DR Repeat.
 KW Repeat.
 FT DOMAIN 519 524 POLY-PRO.
 FT 636 657 GLN-RICH.
 FT DOMAIN 684 688 POLY-PRO.
 FT 699 704 POLY-PRO.
 FT DOMAIN 814 821 POLY-PRO.
 FT 821 821 POLY-PRO.
 FT DOMAIN 1340 1345 POLY-GLY.
 FT 1398 1403 POLY-GLY.
 FT DOMAIN 1436 1442 POLY-PRO.
 FT 1982 1991 POLY-PRO.
 FT DOMAIN 41 1795 4 X 57 AA TYPE A REPEATS.
 FT REPEAT 41 95 1-1.
 FT REPEAT 98 154 1-2.
 FT REPEAT 281 337 1-3.
 FT 1740 1795 1-4.
 FT DOMAIN 337 549 2 X TYPE B REPEATS.
 FT REPEAT 337 418 2-1.
 FT REPEAT 476 549 2-2.
 FT 1899 2089 3 X 50 AA TYPE C REPEATS.
 FT REPEAT 1899 1948 3-1.
 FT REPEAT 1965 2014 3-2.
 FT REPEAT 2040 2089 3-3.
 FT CONFLICT 57 57 R -> A (IN REF. 2).
 FT CONFLICT 109 109 Q -> S (IN REF. 2).
 FT CONFLICT 414 414 P -> PPHRGAPAGMGPP (IN REF. 2).
 FT CONFLICT 532 532 T -> K (IN REF. 2).
 FT CONFLICT 682 682 Q -> K (IN REF. 2).
 FT CONFLICT 730 730 E -> D (IN REF. 2).
 FT CONFLICT 750 750 L -> R (IN REF. 2).
 FT CONFLICT 834 834 G -> A (IN REF. 2).
 FT CONFLICT 1035 1035 M -> L (IN REF. 2).
 FT CONFLICT 1068 1068 P -> R (IN REF. 2).
 FT CONFLICT 1285 1285 G -> A (IN REF. 2).
 FT CONFLICT 1400 1400 T -> S (IN REF. 2).
 FT CONFLICT 1611 1611 G -> A (IN REF. 2).
 FT CONFLICT 1729 1729 G -> A (IN REF. 2).
 SQ SEQUENCE 2142 AA; 227840 MW; 32DDF16B9B52420A CRC64;
 Query Match 11.1%; Score 107; DB 1; Length 2142;
 Best Local Similarity 25.0%; Pred. No. 4.6;
 Matches 46; Conservative 16; Mismatches 48; Indels 74; Gaps 10;
 QY 14 PAPPCVPAACFDLTVHCVACGLRTPRPKAPGASSPAPRTALQPOESVAGAGBALP 73
 Db 506 PAPPAPVPEK-----LAP-PAPPAPAPPTET--EEEPQAPAPAST-P 547
 QY 74 LPELLGAPALGLALVLAIVLVGLVSWRRRRRLRGASSAEPDG-----DK 121

Db 548 TRGVA-AAPLTVS-----GGGSSSTSSGSPFASPVPEQLPSK 584
 QY 122 DAPPELDKXITITSPGISDATAPAPPPEDPCTTPGHSVPVPATLSTELVTTKTAG- 180
 Db 585 EGGEPEPEEV-----PPP-----TTPVPKVEPKGDGIGTPROPQGLGY 624
 QY 181 PEQO 184
 Db 625 PKQO 628
 RESULT 8
 HXB2 HUMAN
 ID HXB2 HUMAN STANDARD; PRT; 356 AA;
 AC P14652; P17485; P10913;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Homeobox protein Hox-B2 (Hox-2H) (Hox-2.8) (K8).
 GN HOXB2 OR HOX2H.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90098876; PubMed=2574852;
 RA Acampora D., D'Esposito M., Fatarella A., Pannese M., Migliaccio E.,
 RA Morelli F., Stornaiuolo A., Nigro V., Simeone A., Boncinelli E.;
 RL "The human HOX gene family.";
 RN Nucleic Acids Res. 17:10385-10402(1989).
 RN [2]
 RP SEQUENCE OF 132-208 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=89378558; PubMed=2570724;
 RA Giampaolo A., Acampora D., Zappavigna V., Pannese M.,
 RA D'Esposito M., Care A., Fatarella A., Stornaiuolo A., Russo G.,
 RA Simeone A., Boncinelli E., Peschle C.;
 RT "Differential expression of human HOX-2 genes along the anterior-
 RT posterior axis in embryonic central nervous system.";
 RL Differentiation 40:191-197(1989).
 RN [3]
 RP SEQUENCE OF 143-208 FROM N.A.
 RX MEDLINE=90215256; PubMed=2576652;
 RA Boncinelli E., Acampora D., Pannese M., D'Esposito M., Somma R.,
 RA Gaudino G., Stornaiuolo A., Cafiero M., Fatarella A., Simeone A.;
 RT "Organization of human class I homeobox genes.";
 RL Genome 31:745-756(1989).
 RN [4]
 RP SEQUENCE OF 143-202 FROM N.A.
 RX MEDLINE=88329001; PubMed=2801346;
 RA Kongswan K., Webb E., Housiaux P., Adams J.M.;
 RT "Expression of multiple homeobox genes within diverse mammalian
 RT haemopoietic lineages.";
 RL EMBO J. 7:2131-2138(1988).
 RN [5]
 RP SEQUENCE OF 1-42 FROM N.A.
 RX MEDLINE=95181447; PubMed=7876223;
 RA Vieille-Grosjean I., Huber P.;
 RT "Transcription factor GATA-1 regulates human HOXB2 gene expression in
 RT erythroid cells.";
 RL J. Biol. Chem. 270:4544-4550(1995).
 CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
 CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
 CC SPECIFIC POSTNATAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN WHOLE EMBRYOS AND FETUSES AT
 CC 5-9 WEEKS FROM CONCEPTION.
 CC -1- SIMILARITY: BELONGS TO THE ANTP HOMEOBOX FAMILY.
 CC PROBOSCIPEDIA SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL, X16665; CAA34655.1; -
DR EMBL, X16176; CAA34298.1; -
DR EMBL, X14571; CAA32709.1; -
DR EMBL, X78978; CAA55581.1; -
DR PIR, S07542; MTHU2H.
DR PIR, E37042; E37042.
DR HSSP, P14653; 1B72.
DR TRANSFAC, T03323; -
DR Genem; HGNC:5113; HOXB2.
DR MIM: 142967; -
DR InterPro; IPR001827; Antennapedia.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00025; ANTENNAPEdia.
DR PRINTS; PR00024; HOMEBOX.
DR Prodom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX 1; 1.
DR PROSITE; PS00032; ANTENNAPEdia; 1.
DR PROSITE; PS00071; HOMEBOX 2; 1.
DR Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Transcription regulation.
FT DOMAIN 94 99 ANTP-TYPE HEXAPEPTIDE.
FT DNA_BIND 143 202 HOMEBOX.
FT CONFLICT 136 138 PEA -> RRL (IN REF. 2).
SQ SEQUENCE 356 AA; 37913 MW; D7ACA0E3D5BEBB9 CRC64;

Query Match 11.0%; Score 106.5; DB 1; Length 356;
Best Local Similarity 24.7%; Pred. No. 0.99; Indels 89; Gaps 12;
Matches 56; Conservative 16; Mismatches 66; Indels 89; Gaps 12;
QY 5 PRS-LRGDAPA-----PTPCVACPCFDLVRHCVAAGLRTTRP-----KPA- 46
DB 66 PRSQRAEGRALPPPPPPPLA-----APAPPEPPMKKKSAKKPSQ 109
QY 47 GASSPAPRTALQPOSSVGAGGAGALPLPG-----LL-----FG 80
DB 110 SATSSPSAPASAVPAGVSGPADGLPEAGGGARLRRTAYNTQLEKEFEHFKYLC 169
QY 81 APALGLALVLAIVLVGLVSW--RRQRRLRGASSSEADGDKDAPPELDKYI-----I 132
DB 170 RRRRVEIALALDLTERQVVFQNRMRKIKRQTOHREPPDGPACGALEDICDPAEBPA 229
QY 133 LSPGISDATAPAM-----PP-----PG-----EDPGTTPG 158
DB 230 ASFGPSASRAAMWACHPPEVVPFALSDPPRLAVRLGAGASSFG 276

RESULT 9
WASL_BOVIN STANDARD; PRT; 505 AA.
ID 095107;
DT 16-OCT-2001 (Rel. 40; Created)
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 15-JUN-2002 (Rel. 41; Last annotation update)
DE Neutral Wiskott-Aldrich syndrome protein (N-WASP).
GN WASL.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97050838; PubMed=8895577;
RA Miki H., Miura K., Takenawa T.;

RT "N-WASP, a novel actin-depolymerizing protein, regulates the cortical
RT cytoskeletal rearrangement in a p12-dependent manner downstream of
RT tyrosine kinases.";
RL EMBL J. 15:5326-5335(1996).

CC -1- FUNCTION: REGULATES ACTIN POLYMERIZATION BY STIMULATING THE ACTIN-
CC NUCLEATING ACTIVITY OF THE ACTIN-RELATED PROTEIN 2/3 (ARP2/3)
CC COMPLEX.
CC -1- SUBUNIT: BINDS ACTIN AND ARP2/3 COMPLEX; INTERACTS WITH CDC42
CC BINDS TO SH3 DOMAINS OF ASH/GRB2.
CC -1- SIMILARITY: CONTAINS 1 CRIB DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 WH1 DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 WH2 DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL, D67066; BA11082.1; -
DR InterPro; IPR000095; PAKbox/RhoGndg.
DR InterPro; IPR000697; RanBP1_WASP.
DR InterPro; IPR001960; WH1.
DR InterPro; IPR003124; WH2.
DR Pfam; PF00568; WH1; 1.
DR Pfam; PF00786; PBD; 1.
DR Pfam; PF02205; WH2; 2.
DR SMART; SM00285; PBD; 1.
DR SMART; SM00461; WH1; 1.
DR SMART; SM00246; WH2; 2.
DR PROSITE; PS50108; CRIB; 1.
KW Actin-binding; Repeat.
FT DOMAIN 34 138 WH1.
FT DOMAIN 203 216 CRIB.
FT DOMAIN 277 392 PRO-RICH.
FT DOMAIN 405 422 WH2 1.
FT DOMAIN 433 450 WH2 2.
FT DOMAIN 486 505 ASP-RICH.
SQ SEQUENCE 505 AA; 54671 MW; 54B83B48F1CB3B8 CRC64;

Query Match 10.7%; Score 103; DB 1; Length 505;
Best Local Similarity 23.6%; Pred. No. 2.3; Indels 84; Gaps 7;
Matches 41; Conservative 4; Mismatches 45; Indels 84; Gaps 7;
QY 5 PRSLRGDAPAPTPC-VPAECFDLVRHCVAAGLRTTRPKPA-----GASSPAPRTALQ 59
DB 301 PPPARGGAPPPPPPSRAPTA-----PPPPPSRGVAGAPPPPPRMP 345
QY 60 QESVAGAGGAGALPLPGILFGAPALGLALVLAIVLVGWSRRQRRLRGASSSEADG 119
DB 346 -----PLPAL-----PSSAPSG 357

RESULT 10
ICPO_HSVBK STANDARD; PRT; 676 AA.
ID P29836;
DT 01-APR-1993 (Rel. 25; Created)
DT 01-APR-1993 (Rel. 25; Last sequence update)
DT 16-OCT-2001 (Rel. 40; Last annotation update)
DE Trans-acting transcriptional protein ICPO (p135 protein) (IER
DE 2.9/IER.6).
GN ICPO.
OS Bovine herpesvirus type 1 (strain K22).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicelloviruses.
OX NCBI_TaxID=31519;

```

RN [1]
RP MEDLINE=92219360; PubMed=1313901;
RX Murch U.V., Friefel C., Vogt B., Vojtek C., Paces V., Schwytzer M.,
RT "Immediate-early RNA 2.9 and early RNA 2.6 of bovine herpesvirus 1
RT are 3' coterminal and encode a putative zinc finger transcriptional
RT protein."
RL J. Virol. 66:2763-2772(1992).
CC -1- PFM: THE STRONGLY ACIDIC REGION MIGHT SERVE AS A TRANSCRIPTIONAL
CC ACTIVATION DOMAIN, POSSIBLY REGULATED THROUGH PHOSPHORYLATION BY
CC CASEIN KINASE II.
CC -1- SIMILARITY: TO OTHER HERPESVIRUSES ICPO PROTEIN.
CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
EMBL, M84464; AAA46061.1; -.
DR F1R; A38209; EDBE22.
DR HSSP; P28990; 1CHC.
DR InterPro; IPR001841; Znf_fing.
DR Pfam; PR00097; ZF-C3HC4; 1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00518; ZF_RING_1; 1.
DR PROSITE; PS50089; ZF_RING_2; 1.
DR Transcription regulation; Trans-acting factor; Activator; Zinc-finger;
DR DNA-binding; Early protein; Repressor; Phosphorylation.
DR ZN FING 13 52 RING-TYPE.
FT DOMAIN 284 331 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 676 AA; 67701 MW; 98B0683C9BFC65D CRC64;

Query Match 10.6%; Score 102.5; DB 1; Length 676;
Best Local Similarity 29.4%; Pred. No. 3.3;
Matches 60; Conservative 14; Mismatches 79; Indels 51; Gaps 11;

QY 7 SURGRAPAPTPVPECEFDLVRHCVAGC-----LIRTRPRAP 45
DB 353 STRGRQTPVQPPAPRS-----LARR--PCGRAAVAPSSRSRGRRPRLPAAFRAP 405
QY 46 AG---ASSPAPRTALPOESVGAGAGALPLPGLLFGAPALLGLALVLVLVG---LV 99
DB 406 AAOARACSPFPR-----EEGRAGLGVAAAGETRGWGVSEEGGERAKILSGAGPPRVQ 460
QY 100 SWRRRQRL-RCASSAEPDGDGDAPEDPLDKVILSPGISDATA-PAMPPGEGDPTTP 157
DB 461 ARRRRTTELDRAPTPAPAP---APAPAPISTMIDL---ANAPARAPDAPAPALGPALA 514
QY 158 GHSVPVPATELGSTELVTTKTAP 181
DB 515 GAOIGTPA-----AAAAVTAAAAP 534

RESULT 11
AREA_PENCH STANDARD; PRT; 725 AA.
ID AREA_PENCH Q01582;
AC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nitrogen regulatory protein area (Nitrogen regulator nre).
GN AREA OR NRE.
OS Penicillium chrysogenum.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX NCBI_Taxid=5076;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95308537; PubMed=7788718;

```

```

RA Haas H., Bauer B., Redl B., Stoeffler G., Marzluft G.A.;
RT "Molecular cloning and analysis of nre, the major nitrogen regulatory
RT gene of Penicillium chrysogenum."
RL Curr. Genet. 27:150-158(1995).
CC -1- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
EMBL, U02612; AA863400.1; -.
DR HSSP; P17429; 4GAT.
DR InterPro; IPR00679; Znf_GATA.
DR Pfam; PR00320; GATA; 1.
DR PRINTS; PR00619; GATAZNFINGER.
DR SMART; SM00401; ZNF_GATA; 1.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
DR PROSITE; PS50114; GATA_ZN_FINGER_2; 1.
DR Transcription regulation; Activator; DNA-binding; Zinc-finger;
DR Nuclear protein; Nitrate assimilation.
FT ZN FING 525 549 GATA-TYPE.
SQ SEQUENCE 725 AA; 76848 MW; 45C24148F1FAD43 CRC64;

Query Match 10.6%; Score 102.5; DB 1; Length 725;
Best Local Similarity 27.0%; Pred. No. 3.5;
Matches 47; Conservative 13; Mismatches 59; Indels 55; Gaps 7;

QY 32 CVACGL-----LIRTRPRAGASP-----APRPL 57
DB 546 CNACGLFLKGVNPLSLKTDVYIKKRRSSANLTVGTSSSKSKSRKNSIQHAPSTSI 605
QY 58 QPQ-----ESVGAGAGALPLPGL--LFGAPALLGLALVLVLGLVSWRRQRRLGA 111
DB 606 SSRNMTSEPPINSSTLKGQGVPIAAPPKSPGAGVAGARQGVAPARQRLLEA 665
QY 112 SGAEPDGDKD---APEPLDKVILSPGISDATA-PAMPPGEGDPTTPGHV 161
DB 666 PAGESPDADDSPEKSAAPPSRSKVPL-----APAMAPPA---AANPANHSI 708

RESULT 12
YQ35_CAEEL STANDARD; PRT; 317 AA.
ID YQ35_CAEEL Q09456;
AC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Putative cuticle collagen C0965.5.
GN C0965.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Br1etol N2;
RA Palmer S.;
RL Submitted (NOV-1994) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT (BY SIMILARITY).
CC -1- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE
CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
CC LINKS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CUTICULAR COLLAGEN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```


CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; 246791; CAA6758.1; -
 DR Wormpep; C09G5.5; CE01485.
 DR InterPro; IPR002486; Col cuticle_N.
 DR InterPro; IPR000087; Collagen.
 DR Pfam; PF01391; Collagen; 3.
 DR Pfam; PF0184; Col cuticle_N; 1.
 DR Hypothetical protein; Cuticle; Connective tissue; Repeat;
 KW Multigene family; Collagen.
 FT DOMAIN 92 124 TRIPLE-HELICAL REGION.
 FT DOMAIN 137 199 TRIPLE-HELICAL REGION.
 FT DOMAIN 202 264 TRIPLE-HELICAL REGION.
 FT SEQUENCE 317 AA; 31283 MW; 685DCF24612707BB CRC64;

Query Match 10.6%; Score 102; DB 1; Length 317;
 Best Local Similarity 26.7%; Pred. No. 1.8;
 Matches 54; Conservative 10; Mismatches 68; Indels 70; Gaps 12;

QY 4 GRSLSRGDAPAPTPCPV--AACEFDLVHVCVACGLLRTPR--PKDAGA-SSPAPR---- 54
 DB 101 GDRGLDQGPAGKPGQPGVAGPAHHQOQECIKC-----PGAGPAPGAPGNPQGPNG 155
 QY 55 TLQLOQESVAG-----AGEAALP---LPGILFGAPALLGLALVLAIVLVGLVSWRR 103
 DB 156 NGAAPAHGGGGPPEGPAGAGGAGGAGAPGNP--GRPGGSG-----Q 198
 QY 104 RQRRLRGASSAAPDDKDAPEPLDKVILSPGISDATAPAMP-----PGEPP 152
 DB 199 RRRGLPGEGRPGCPGPAPG-----QPGSSGTPGPAGPPGPPGNGGPHGQNG 250
 QY 153 GTTPGHS-----VEVPA 165
 DB 251 QGAPGNDGAPGSDAAYCPCPA 272

RESULT 13
 DBP_MOUSE STANDARD; PRT; 325 AA.
 ID DBP_MOUSE
 AC 060925;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 D-site-binding protein (Albumin D box-binding protein).
 GN DBP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCB1_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/Sv; TISSUE=Liver;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 RL - FUNCTION: THIS TRANSCRIPTIONAL ACTIVATOR RECOGNIZES AND BINDS TO
 - THE SEQUENCE 5'-RTTAGTAA-3' FOUND IN THE PROMOTER OF GENES SUCH
 - AS ALBUMIN, CYP2A4 AND CYP2A5. IT IS NOT ESSENTIAL FOR CIRCADIAN
 - RHYTHM GENERATION, BUT MODULATES IMPORTANT CLOCK OUTPUT GENES. MAY
 - BE A DIRECT TARGET FOR REGULATION BY THE CIRCADIAN PACEMAKER
 - COMPONENT CLOCK. MAY AFFECT CIRCADIAN PERIOD AND SLEEP REGULATION
 - (BY SIMILARITY).
 - SUBUNIT: BINDS DNA AS A HOMODIMER OR A HETERODIMER. CAN FORM A
 - HETERODIMER WITH TEF (BY SIMILARITY).
 - TISSUE SPECIFICITY: EXPRESSED IN THE SUPRACHIASMATIC NUCLEI (SCN)
 - AND IN MOST PERIPHERAL TISSUES, WITH A STRONG CIRCADIAN
 - RHYTHMICITY.
 - SIMILARITY: BELONGS TO THE BZIP FAMILY. PAR SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; U29762; AAA73924.1;
 DR MGD; MGI:94866; DBP.
 DR InterPro; IPR004827; TP_bZIP.
 DR SMART; SM00338; BRLZ; 1.
 DR DNA-binding; Transcription regulation; Activator; Nuclear protein;
 KW Biological rhythms.
 FT DOMAIN 129 135 POLY-PRO.
 FT DOMAIN 156 161 POLY-SER.
 FT DOMAIN 188 239 PRO-RICH (PROLINE/ACIDIC REGION (PAR)).
 FT DNA BIND 258 277 BASIC MOTIF.
 FT DOMAIN 290 304 LEUCINE-ZIPPER.
 FT SEQUENCE 325 AA; 34428 MW; 58F7067DD1238B06 CRC64;

Query Match 10.6%; Score 102; DB 1; Length 325;
 Best Local Similarity 26.1%; Pred. No. 1.8;
 Matches 54; Conservative 14; Mismatches 77; Indels 62; Gaps 9;

QY 4 GRSLSRGDAPAPTPCPV--AACEFDLVHVCVACGLLRTPR--PKDAGA-SSPAPR---- 54
 DB 83 GGGSPRSGSGVAGSLPAPLIMERTPPGQVEYVLAFLER---GL--PPSPPPG 136
 QY 48 ASSPAPRTALQOESVAGAGEAALP--LPGILFGAPALLGLALVLAIVLVGLVSWRRQ 105
 DB 137 GLSPAPSPAPRTAPSPGSGSSSSPSRSPG---HAP-----R 172
 QY 106 RRLRGASSAAPDDKDAPEPLDKVILSPGISDATAPAMP-----PGEPP 152
 DB 173 ATLGAAGHRAGLTSRDPSPVD-----PIVEVLTMTFEPDPAALASIPGHETFDPR 226
 QY 166 TELGSTELY-----TTKTAGEPQQ 184
 DB 227 RHRFEELKQPIKKAKKQVPEEQ 253

RESULT 14
 DBP_RAT STANDARD; PRT; 325 AA.
 ID DBP_RAT
 AC P16443;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 D-site-binding protein (Albumin D box-binding protein) (D site albumin
 DE promoter binding protein 1).
 GN DBP.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCB1_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WiStar; TISSUE=Liver;
 RL MEDLINE=90235277; PubMed=2331750;
 RL "DBP, a liver-enriched transcriptional activator, is expressed late
 RL in ontogeny and its tissue specificity is determined
 RL posttranscriptionally."
 RL Cell 61:279-291 (1990).
 RN [2]
 RP REVISIONS.
 RA Mueller C.R., Mairre P., Schibler U.;
 RL Cell 65:915-915 (1991).
 RN [3]
 RP REVIEW.
 RX MEDLINE=9439627; PubMed=10508692;

RA Brown S.A., Schibler U.;
 RT "The ins and outs of circadian timekeeping."
 RL Curr. Opin. Genet. Dev. 9:588-594(1999).
 CC -1- FUNCTION: THIS TRANSCRIPTIONAL ACTIVATOR RECOGNIZES AND BINDS TO
 CC THE SEQUENCE 5'-ATTAGTAA-3' FOUND IN THE PROMOTER OF GENES SUCH
 CC AS ALBUMIN, CYP2A4 AND CYP2A5. IT IS NOT ESSENTIAL FOR CIRCADIAN
 CC RHYTHM GENERATION, BUT MODULATES IMPORTANT CLOCK OUTPUT GENES.
 CC BE A DIRECT TARGET FOR REGULATION BY THE CIRCADIAN PACEMAKER
 CC COMPONENT CLOCK. MAY AFFECT CIRCADIAN PERIOD AND SLEEP REGULATION
 CC (BY SIMILARITY).
 CC SUBUNIT: BINDS DNA AS A HOMODIMER OR A HETERODIMER. CAN FORM A
 CC HETERODIMER WITH TEF.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE SUPRACHIASMATIC NUCLEI (SCN)
 CC AND IN MOST PERIPHERAL TISSUES, WITH A STRONG CIRCADIAN
 CC RHYTHMICITY.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED LATE IN ONTOGENY.
 CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY. PAR SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: J03179; AAA41083.1; -
 CC PIR: A34894; A34894.
 CC PIR: A39531; A39531.
 CC TRASNFA: T00183; -
 CC InterPro: IPR004827; TF_BZIP.
 CC SMART: SM00338; BRLZ; 1.
 CC DR DNA-binding; Transcription regulation; Activator; Nuclear protein;
 CC KW Biological rhythm; KM
 FT DOMAIN 129 135 POLY-PRO.
 FT 161
 FT DOMAIN 188 239 POLY-SER.
 FT 258 277 PRO-RICH (PROLINE/ACIDIC REGION (PAR)).
 FT DOMAIN 290 304 BASIC MOTIF.
 FT SEQUENCE 325 AA; 34436 MW; D9B2A53FF18455B2 CRC64;
 SO
 Query Match 10.6%; Score 102; DB 1; Length 325;
 Best Local Similarity 26.1%; Pred. No. 1.8;
 Matches 54; Conservative 14; Mismatches 77; Indels 62; Gaps 9;
 QY 4 GRRSLRGRRAPARTPCVPA-----RCFPL---LVNHCACAGLRTPRKPKAG 47
 83 GGSGPRGRGPAVAGPSLFAPLWERTLPQGVYVLDLAFLEH---GL--PPSPPPG 136
 48 ASSPARTRALQPOESVAGAGEALP--LFGLLFGAPALLGLALVLAIVGLVSWRRQ 105
 137 GLSPASPARTAPASGPGSCSSSRSSPG---HAPA-----R 172
 QY 106 RLRLGASSAEPDGDYDAPEPLDKVILSPGISDATAPAMPPEGDEPTTPGSHVAPA 165
 173 ATLGAAGRHAGLTSRDTSPVD-----PDVVEVLMTEPEPDADLALSLPCHETFDPR 226
 QY 166 TELGSTEIV-----TTKTAGPEEQ 184
 DB 227 RHRFSEELKPOPIMKKARKVQVEEQ 253
 RESULT 15
 IIR7_HUMAN
 ID IIR7_HUMAN STANDARD; PRT; 503 AA.
 AC Q92985; Q00331; Q00332; Q00333; Q75924;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Interferon regulatory factor 7 (IRF-7).
 GN IIR7.
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_Taxid:9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM A).
 RC TISSUE=spleen;
 RA Grossman A., Nicholl J., Antonio L., Luethy R., Suggs S.,
 RA Sutcliffe G.R., Mak T.W.;
 RL Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS A; B AND C).
 RX MEDLINE=97459673; PubMed=9315633;
 RA Zhang L., Pagano J.S.;
 RT "IRF-7, a new interferon regulatory factor associated with Epstein-
 RT Barr virus latency."
 RL Mol. Cell. Biol. 17:5748-5757(1997).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM D).
 RX MEDLINE=99003279; PubMed=9786932;
 RA Au W.-C., Moore P.A., Lafleur D.W., Lafleur P.M.;
 RT "Characterization of the interferon regulatory factor-7 and its
 RT potential role in the transcription activation of interferon A
 RT genes".
 RL J. Biol. Chem. 273:29210-29217(1998).
 CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR. BINDS TO THE INTERFERON-
 CC STIMULATED RESPONSE ELEMENT (ISRE) IN IFN PROMOTERS AND IN THE Q
 CC PROMOTER (QP) OF EBV NUCLEAR ANTIGEN-1 (EBNA1).
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS: A (SHOWN HERE), B/BETA, C/GAMMA
 CC AND D/H; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN SPLEEN, THYMUS, AND
 CC PERIPHERAL BLOOD LEUKOCYTES.
 CC -1- SIMILARITY: BELONGS TO THE IRF FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U73036; AAB17190.1; -
 CC EMBL: U53830; AAB80686.1; -
 CC EMBL: U53831; AAB80688.1; -
 CC EMBL: U53832; AAB80690.1; -
 CC EMBL: AF076494; AAC70999.1; -
 CC HSP: P23906; ZIRF.
 CC TRASNFA: T04674; -
 CC GeneW; HGNC:6122; IRF7.
 CC MIM: 605047; -
 CC InterPro: IPR001346; IRF.
 CC Pfam: PF00605; IRF.1.
 CC PRINTS: PR00267; INTERFERON.
 CC ProDom: PD002355; IRF.1.
 CC SMART: SM00348; IRF.1.
 CC PROSITE: PS00601; IRF.1.
 CC DR PROSITE; PS00601; IRF.1.
 CC Transcription regulation; DNA-binding; Nuclear protein; Activator;
 CC KW Alternative splicing;
 FT DNA_BIND 13 122
 FT VARSPLIC 1 6
 FT VARSPLIC 152 164
 FT VARSPLIC 165 503
 FT VARSPLIC 228 256
 FT CONFLICT 179 179 MISSING (IN ISOFORM B).
 FT CONFLICT 412 412 E -> K (IN REF. 2).
 FT SEQUENCE 503 AA; 54278 MW; AAG39E0E272727C CRC64;
 SO
 Query Match 10.3%; Score 99.5; DB 1; Length 503;
 Best Local Similarity 29.0%; Pred. No. 4;
 Matches 42; Conservative 8; Mismatches 68; Indels 27; Gaps 5;

```

Oy 10 GRDAPAPTECVPAECFDLIV---RHCVACGLIRTPRPKAGASSPAPRTALQPCEVGA 65
Db 165 GLQAFQPLPAPAGDEGDLILQAVQSCCLADHLLTA-----SWGADVPPTKA-----P 211
Oy 66 GAGEAALPPLGILLFGAPALLGLALVLAIVLVGIVSVRRRORRLRGASSAEPDGDKAPE 125
Db 212 GEGQEGPLPLTGACACAGP-----GLPAGELYGWAVEITPSPGPQPAALTTGEAAPE 262
Oy 126 -PLDKVILSPGISDATAPAMPPEG 149
Db 263 SPHQAEPLYLSPSPSACTAVQEPSPG 287

```

Search completed: June 23, 2003, 15:14:13
 Job time : 25 secs

THIS PAGE BLANK (USPTO)